DOCUMENT RESUME

ED 384 613 TM 023 199

AUTHOR Cizek, Gregory J.; And Others
TITLE Further Investigation of Teachers' Assessment Practices.
PUB DATE Apr 95
NOTE 43p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Assessment; Educational Policy; *Educational Practices; *Elementary School Teachers; Elementary Secondary Education; Experience; Grades (Scholastic); *Grading; School Districts; *Secondary School Teachers; Sex Differences; Student Evaluation; Surveys; *Teacher Attitudes; Teacher Characteristics; Test Construction; *Testing

ABSTRACT

A sample of 143 midwestern elementary and secondary school teachers from a variety of practice settings responded to a survey and provided comments regarding their assessment and practices. The study collected background (demographic) information on the teachers and information on some assessment-related practices, including: (1) the frequency with which teachers assign routine class assignments; (2) the types of marks used to report student performance; (3) the frequency and grading of major assignments and tests; (4) the source of classroom tests; (5) the kinds of marks used; (6) the methods used to combine marks; (7) the meaning of grades; (8) teachers' knowledge and perceptions regarding district grading policies; and (9) teachers' awareness of the grading policies of their peers. It was found that assessment practices vary widely and unpredictably. Few relationships were observed between teachers' assessment and grading practices and personal or background characteristics such as practice level, years of experience, gender, or familiarity with district policies. Teachers generally claimed to use a variety of factors in assigning grades, and a majority of teachers surveyed indicated that they were unaware of both their districts' policies and their colleagues' practices. Conclusions, recommendations, and implications of these findings are discussed. (Contains 33 references and 4 tables.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Further Investigation of Teachers' Assessment Practices

ED 384 613

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
GREGORY J. CIZEK

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Gregory J. Cizek
Assistant Professor of Educational Research and Measurement
350 Snyder Hall
University of Toledo
Toledo, OH 43606-3390
Telephone: 419-537-2611

Robert E. Rachor
Graduate Assistant
375 Snyder Hall
University of Toledo
Toledo, OH 43606
419-537-4615

Shawn Fitzgerald
Graduate Assistant
375 Snyder Hall
University of Toledo
Toledo, OH 43606
419-537-4615

Paper presented at the annual meeting of the American Educational Research Association, April 1995, San Francisco, CA.

BEST COPY AVAILABLE

ABSTRACT

A sample of 143 midwestern elementary and secondary school teachers from a variety of practice settings responded to a survey and provided comments regarding their assessment and practices. The survey collected background (demographic) information on the teachers, and information on several assessment-related practices, including: 1) the frequency with which teachers' assign routine class assignments; 2) the types of marks used to report student performance; 3) the frequency and grading of major assignments and tests; 4) the source of classroom tests; 5) the kinds of marks used; 6) the methods used to combine marks; 7) the meaning of grades; 8) teachers' knowledge and perceptions regarding district grading policies; and 9) teachers' awareness of the grading policies of their peers. Interviews with the teachers provided additional insights into their practices.

It was found that assessment practices vary widely and unpredictably. Few relationships were observed between teachers' assessment and grading practices and personal or background characteristics such as practice level, years of experience, gender, or familiarity with district policies. Teachers generally claim to consider and incorporate a variety of diverse factors in assigning grades, and a majority of the teachers surveyed indicated that they were unaware of both their districts' policies and their colleagues' practices. Conclusions, recommendations, and implications of these findings are discussed.

Further Investigation of Teachers' Assessment Practices

Much of the recent renewed interest in educational assessment has been targeted toward two aspects: 1) large-scale testing and its uses and influences on teaching and learning, and 2) investigations of alternate assessment formats. These concerns are related: They both focus on information gathering. As Airasian (1994) has argued, nearly all of the assessment-related activities that teachers engage in can be broadly conceived as information gathering. Airasian defines assessment as "the process of collecting, synthesizing, and interpreting information to aid in [educational] decision making" (1994, p. 5).

In contrast to the recent attention to information gathering, comparatively little attention has been given to information reporting. This component is exemplified by the assigning of grades, marks, or summative evaluations of student performance.

Background

Investigation of teachers' assessment practices has a long history. Early research in the area by Starch and Elliott (1912, 1913a, 1913b) addressed the (un)reliability of grades teachers assigned in high school courses such as English, history, and math; the research program in these areas continues today (cf. Stiggins, Frisbie, & Griswold, 1989).

A number of reviews of research have summarized what is known about the meaning of teacher-based judgments of student achievement (see Brookhart, 1994; Hoge & Coladarci, 1989) and the discouraging picture presented in the first part of the century has apparently not changed appreciably. For example, in a recent review that found "a gap between current practice and measurement theory," Brookhart concluded that teachers' grading practices often "confound constructs into composite scores of questionable reliability and validity and thus uncertain meaning" (1994, p. 299).

Previous research has provided some explanations for the weaknesses in classroom assessment. For example, Hills (1991) has documented a general lack of interest in testing and grading on the part of teachers. A research program conducted by Stiggins and his associates has investigated the nature of teachers' assessment practices and has stressed the importance of knowledge about assessment as a key to educational reform (see Stiggins, 1988; 1991a; Stiggins & Bridgeford, 1985). Despite its seemingly obvious relevance to teachers' practice, it also appears that teachers need assistance in acquiring knowledge about sound assessment. Various researchers have demonstrated the lack of training in educational measurement (see Gullickson, 1986; Schafer & Lissitz, 1987; Ward, 1980; Wise, Lukin, & Roos, 1991); others have suggested ways of conceptualizing and providing assessment training so that it is more relevant to teachers' classroom assessment needs (see Airasian, 1991; Stiggins, 1991b; Stiggins & Conklin, 1988).

Another area of research has focussed on grades and the content of school report cards. An interesting history of grading at several Ivy League schools is provided by Durm (1993). An analysis of current content of report cards has been reported by Gorney (1993).

Friedman and Frisbie (1993) examined 216 report cards used in kindergartens, elementary schools, middle schools, and high schools in Wisconsin; as in previous studies, their work raised questions about the validity of report cards as indicators of student performance. In a companion study, Waltman and Frisbie (1993) studied how the content of report cards facilitates or hinders parents' understanding of the information provided in the reports. They concluded that, if report cards are viewed as a vehicle for accomplishing a transmission of a teacher's intended meaning to parents, the job is not generally being accomplished successfully.

The present research focuses on the grading practices at the level of the individual teacher, and joins an ongoing line of research in that area. Previous work has investigated:

- preservice teachers' beliefs about grading (Jones, 1990);
- the meaning of grades assigned by teachers (Brookhart, 1993);
- the effect of various student characteristics on the grades inservice teachers assign (Griswold & Griswold, 1992; Manke & Loyd, 1990; Nava & Loyd, 1992; Wood, Bennett, Wood, & Bennett, 1990); and
- developing conceptualizations of, systematic approaches to, and a research agenda for grading (Stiggins, Frisbie, & Griswold, 1989; Terwilliger, 1989).

Objectives

The research reported in this paper attempts to provide additional insights into the assessment practices in elementary and secondary schools. Specifically, this paper presents:

1) a description of some current assessment and assessment reporting practices of inservice teachers; 2) an investigation into the possible existence of differences in assessment and assessment reporting practices based on various personal characteristics and practice settings; 3) a description of some of the elements teachers report to be contained in the grades they assign; 4) an evaluation of the meaning of teachers' grades in light of their intentions; 5) an assessment of the linkages between teachers' assessment practices and school district policies; and 6) an attempt to develop a model which might predict teacher assessment behavior. This paper extends earlier work on the same topics by Cizek and Rachor (1994).

Sample and Procedures

Over one academic year, a survey was administered 143 students at the beginning of an introductory master's level course in measurement and evaluation at an urban midwestern university. Because the course is required for continuing certification, the sample was quite diverse. Table 1 provides a summary of background characteristics of the survey respondents.

Insert Table 1 about here

As Table 1 shows, the sample consisted of 31 males (21.7%) and 112 females (78.3%). This distribution is similar to that reported in other studies of elementary and secondary school teachers. Reported ages ranged from 21 to 63 years ($\bar{x}$ = 39.6, S = 9.6). The sample was also diverse in terms of teaching experience: The two largest groups included

teachers with one to five years of experience (27.3%) and more than 15 years of experience (36.4%). The average years of experience in education was 13.2 years (S = 8.6 years). Fifty-eight percent of the respondents were elementary school teachers, 23.1% were high school teachers, and 18.9% of the sample reported teaching in a middle school or junior high school setting. The school types represented included urban (40.6%), suburban (33.6%), rural (14.7%) and Other/Missing (11.2%).

The course from which respondents were drawn presents introductory concepts and principles of educational measurement; the instrument was administered during the first two weeks of the course to prevent the possibility that responses to survey items would be influenced by the course content. Some of the surveys were completed independently; in other instances the survey was administered in a one-on-one situation, which permitted respondents to ask questions about the survey and to provide the interviewer with additional information or elaborations of their responses.

In addition to the information collected on respondents' background characteristics, the instrument contained items which addressed the respondent's assessment and assessment reporting practices. Questions included items asking about:

- What factors teachers considered when assigning grades on assignments, tests, etc. (e.g., number right, difficulty of the assessment, the performance of the class as a whole, individual student ability, and student effort);
- What the final grade for a marking period represented;
- What sources of information the teacher used in assigning the final grade

(e.g., formal and informal achievement and achievement-related measures);

- The frequency with which the teacher gave minor assignments, and minor, and major tests;
- The sources teachers used to obtain assessments (e.g., developed by self or provided by a publisher);
- The number of total marks the teacher included when calculating a students' final grade;
- The teacher's knowledge of other teachers' grading practices; and
- The teacher's knowledge about relevant district policies on grading.

Results

This section provides results from both the survey and teacher interviews. First, results of the quantitative analysis are presented. Second, we attempt to synthesize what we learned from talking with teachers about their assessment practices.

I. Quantitative Analyses

In this section, we provide results for the analysis of teachers' responses to the survey questions. In the accompanying tables, frequency data is provided for the total group and broken down by practice setting (e.g, elementary, middle school, high school), by gender (male, female), and for years of teaching experience (1-5, 6-10, 11-15, 16+). For each question, a chi-square test of independence was performed to investigate potential relationships between responses and these background variables. Finally, a logistic regression

procedure was used to evaluate the relative contributions of various predictors of teachers' grading philosophies and their knowledge of district grading policies.

Frequency and source of assessments

Table 2a summarizes respondents' answers to two questions about the frequency and sources of classroom assessments, with responses broken down by practice setting (i.e., elementary, middle, and high school). The table shows that about three-fourths (75.2%) of the teachers indicated that they gave minor classroom assignments that counted for a grade at least weekly. The remaining teachers reported giving minor assignments every two weeks or even less often. There were no statistically significant differences in the frequency of minor classroom assignments between elementary, middle school, or high school teachers. Respondents' answers to the question about the frequency of formal tests and exams that count for a grade are also summarized in Table 2a. A slight majority of teachers (53.8%) reported giving major tests about once every two weeks, with the rest giving major tests less frequently. There were no statistically significant differences between elementary, middle school, or high school teachers on this question.

The teachers were also asked about their primary source for minor and major assessments. A majority of the teachers (84.6%) indicated that they usually developed their own minor tests, while 74.2% indicated they usually developed their own major tests and/or examinations. Publishers were their primary source for minor tests or quizzes for 15.4% of the teachers and 25.8% indicated that publishers were their primary source for major tests. The calculated chi-square for the question on sources of minor tests between elementary,

middle school, and high school teachers was statistically significant ($\chi^2 = 6.50$, df = 2; p = .03). However all standardized residuals were less than 2.00 (in absolute value), indicating that, comparatively, no group was a major contributor to the significant chi-square value.

Table 2b provides a summary of the frequency and sources of assessments with responses broken down by the teachers' years of experience. For this analysis, teachers were grouped into four categories according to total years of experience in education (1 = one to five years; 2 = six to 10 years; 3 = 11 to 15 years; 4 = 16 or more years). There were no statistically significant differences between these groups in their frequency of giving routine assignments, frequency of major tests, or primary source of major tests. There was, however, a statistically significant difference, based on years of experience, for the teachers' primary source of minor tests and quizzes ($\chi^2 = 9.06$, df = 3, p = .03), with nearly all of the beginning teachers (96.9%) indicating that they develop minor tests and quizzes themselves.

Respondents' answers to the same questions, analyzed by gender, are presented in Table 2c. There were no statistically significant difference between male and female teachers in the frequency of minor assignments, major test, or source for minor tests. There was a statistically significant difference between males and females in their primary source of major tests ($\chi^2 = 6.15$, df = 1, p = .01), with 81.5% of male teachers indicating that they develop their own major tests, and only 54.8% of the female teachers doing so.

Insert Tables 2a, 2b, 2c about here

Factors considered in assigning grades

Respondents were asked which factor or factors they considered when assigning grades to students' assignments, tests, and for final grades. Tables 3a through 3c reveal that teachers considered many factors. Most teachers (83.8%) considered the percentage or number correct on the assignment or test when assigning a grade. Other considerations included the individual student's ability (51.5%), the performance of the entire class on the assignment or test (43.4%), the student's effort on the assignment or test (41.9%), and the difficulty of the assignment or test (35.3%). There were no statistically significant differences between practice settings (elementary, middle, high school), years of experience, or gender in the factors considered when respondents assigned grades to assignments or tests.

There were two questions about final grades for a marking period. In their responses to the first question about sources of information used to assign final grades, teachers indicated that they considered several sources of information. As shown in Table 3a through 3c, most teachers (89%) used formal achievement measures such as tests, assignments, reports, and quizzes. Over half the teachers (52.2%) used other formal, achievement-related measures such as attendance and class participation; 41.9% of the teachers used informal measures of achievement, such as students answers to questions during class or their contributions to discussions; and 61.0% percent used informal, non-achievement related measures such as student conduct, and impressions of a student's effort and teamwork in class when assigning the final grade. Statistically significant differences were found for the use of formal non-achievement related information for elementary, middle school, and high school

teachers (χ^2 = 6.51, df = 2; p = .03). Again, however, all standardized residuals were less than 2.00.

The teachers in this study were also asked what the final grades they assign for a marking period represent. Most teachers seemed to relate final grades, in some manner, to achievement on fixed classroom goals. For example, 34.9% of all teachers indicated that the final grade represented students' individual achievement on fixed classroom goals or objectives; 18.6% indicated that the final grade represented individual achievement on fixed classroom goals, but taking into account the performance of the class as a whole; and 28.7% said that their final grades represented a combination of group and individual work toward fixed classroom goals. The remaining teachers, 17.8% indicated that their final grades represented individual student achievement on individual student goals. There were no statistically significant differences between elementary, middle school, or high school teachers, between genders, or between experience levels.

Insert Tables 3a, 3b, 3c about here

Number of grades assigned and grading policies

Information on the number of grades used to calculate final grades for the semester and respondents' knowledge of the number of grades other teachers used in calculating their final grades is presented in Tables 4a through 4c. On average, respondents used 24.3 (S =

17.9) grades per marking period when calculating students' final grades. There were no statistically significant differences between the number of grades used to calculate the final grade between elementary, middle school, and high school teachers, between male and female teachers, or between levels of teaching experience.

Over half of the respondents (52.7%) did not know how the total number of grades they used to calculate students' final grades compared with other teachers in their building. Twenty-four percent of the teachers indicated that most teachers in their building used about the same number of grades, while 17.8% indicated that other teachers used fewer grades and 5.4% thought other teachers used more.

A slight majority of teachers (55.6%) indicated that their district had a formal grading policy, while approximately one-third (32.6%) indicated that their district did not have such a policy. The rest of the teachers (11.9%) were not sure whether their district had a policy or not. No significant differences for these questions were found for practice setting, gender, or years of experience.

Insert Tables 4a, 4b, 4c about here

Predictors of grading philosophy and knowledge of grading policies

Data from the questionnaires were also analyzed using logistic regression. Three separate analyses were conducted. In the first analysis, the outcome was a dichotomous variable indicating whether or not the teacher reported having knowledge of a district policy

on grading. In the second and third analyses, the criterion was a dichotomous variable indicating whether or not the teacher engaged in the practice of assigning grades based upon fixed individual (second analysis) or fixed individual and fixed class goals (third analysis). Background and practice variables believed to have some relationship to changes in professional practice were included as predictors using a forced entry procedure.

Results for the first analysis (knowledge of district grading policy) indicated that none of the predictor variable coefficients were significantly different from zero. In the second analysis, only one of the predictor variables--number of grades assigned per marking period--was found to have a statistically significant coefficient ($\beta = -.0447$, $df = 1$, $\Psi = .9563$, $p < .05$). Although this result is statistically significant, the substantive interpretation of the result is nearly trivial: The odds ratio indicates that teachers who assign more grades per marking period are 1.04 times more likely to assign grades that do not reflect individual achievement on fixed goals. (XXXShawn: please verify that this is the correct interpretation and not the opposite.)

In the third analysis, the background variable "Years of Experience" was found to have a statistically significant coefficient ($\beta = .3161$, $df = 1$, $\Psi = 1.14$, $p < .05$). In this case, the odds ratio indicates that teachers with more experience are 1.14 times more likely to assign grades that reflect individual achievement on fixed class or individual goals.

II. Analysis of Teacher Comments

The quantitative analysis reported earlier in this paper presented a picture of widely variable assessment practices and knowledge of both measurement fundamentals and district

policies. Analysis of teachers' comments provided some insights into why such variability occurs and some glimpses into the "assessment mindset" of the teachers studied. First, one caveat is warranted. The comments provided by teachers were not solicited; that is, the only comments we recorded were volunteered by the teachers surveyed. We do not claim that these comments represent the thinking of all teachers, or even of all the teachers in the study. On the other hand, they do represent the perspectives of teachers who wanted to provide input on assessment issues.

The comments teachers provided could be classified in many ways. For this analysis, the comments were classified according to three dimensions. First, teachers commented on why they perform assessments, especially formal assessments such as tests and graded assignments. These comments tend to reveal something about the assumptions the teachers bring to the assessment process. Teachers also commented on the targets of their assessment practices and how they use assessment information. These comments help to illustrate what ultimately their grades are based upon and how they might be interpreted. Finally, teachers provided some comments regarding professional collaboration. This collaboration took two forms: a "local" collaboration, as in interactions with other teachers, and a more macro level of collaboration, viewed as the extent of their connection to other levels of professional practice, such as with administrators, or alignment with district-level policies.

Assessment assumptions

As revealed in the section providing some quantitative results, there is remarkable diversity in teachers' assessment perspectives and practices. Some of the teachers' comments

helped to illustrate why that diversity could occur in such apparent harmony. Other comments dealt with assumptions about the accuracy of assessment information. To illustrate that different assessment practices can peacefully coexist, one teacher provided an extended example of why diversity in teachers' assessment practices are, in essence, irrelevant.

> "Different teachers use different methods of measuring progress, but they measure the same behaviors, so, ultimately, the results are similar. For example, a Canadian weatherman uses a Celsius thermometer to measure outdoor temperature while an American weatherman uses a Fahrenheit thermometer, but they both conclude that the temperature is colder than normal for that date."

As this example illustrates, if teachers are all really interested in assessing the same things, (and assuming that the various thermometers are equally accurate) then the fact that teachers choose different assessment methods would have little bearing on the outcome of the process.

Another common assumption was related to a "success orientation" that many teachers possessed. Several teachers described a desire "to give every student every chance to be successful." An underlying assumption seemed to be that, if a student was not successful, then there may be some defect in the thermometer. Some of the comments also seemed to suggest that a teacher's informal assessment of a student based upon many observations would usually be more accurate than the data resulting from a lesser number of more formal

assessment, or on any formal assessment. One teacher provided a comment on how other teachers assess:

> "To measure these students properly, the teachers do not put an overreliance on standard tests. They also include observations to help in developing a more accurate picture of the student and to get a better understanding of the student's abilities. This approach certainly benefits the student and provides a more conclusive evaluation of the student's academic performance."

Finally, we note the comment of one teacher regarding an assumption of the inherent differences in sound measurement practice in public and private schools. This teacher--a teacher in a private school--observed that:

> "Teaching in a private school system requires a more rigid grading policy. One policy is homework. When a teacher at [school name] gives a homework assignment, the teacher is required to give credit for the assignment. The policy is to collect it, grade it, and record it."

Assessment targets and grading

Teachers' comments in the areas of assessment targets and grading again complemented the findings of the quantitative analyses. Most teachers recognized that there are many valuable educational outcomes worthy of assessing; most teachers seemed willing to

try to assess all of them and to form a composite index--the final grade.

Knowledge and skill acquisition were targets teachers assessed, although non-cognitive outcomes were also at least informally assessed and combined to arrive at a grade. In particular, several teachers described the desire to have cooperation or group efforts included as a valuable outcome that should contribute to the final grade. However, it was often unclear whether cooperation was really assessed in the methods used, or how the assessment of cooperation contributed to the final grade. One industrial arts teacher who apparently wanted to clearly delineate the contribution of each component, used a "points" system, and awarded five points each day to each student who cleaned up the work area: Sometimes each student performed the clean up independently, or "When the task is done in pairs or groups of students, in which event, each student would earn five points for cooperative or group work."

Again, although cognitive development was acknowledged as important contributor to students' grades, it was mentioned only rarely, and in passing. We suspect that this occurred because most teachers assumed that assessment of cognitive outcomes is the norm. The teachers' comments were most likely to emphasize the assessment of non-cognitive outcomes. However, some teachers expressed a clear preference for non-cognitive outcomes; as one (elementary) teacher said, "Getting the child through the level with a positive attitude and good memories is more important than a raw number grade... Shaping the kids' minds through group interaction, effort, and participation is more important than averaging tests and quiz scores." Another teacher reported that "assignments, quizzes, and tests are not crucial in [her] grading policies." This teacher "stresses group interaction and uses several other

subjective methods combined with intuition to formulate a final grade."

Perhaps the most extreme example we found related to emphasis on non-cognitive outcomes was the practice of one secondary-level teacher who felt obligated to give tests and quizzes, but who reported that she usually gives the test and the answers to her students to take home and complete.

Attendance and participation were fairly highly valued by the teachers in the sample, as were some of the other non-cognitive outcomes described above. It was particularly interesting to us, however, to learn how teachers reported combining the divergent sources of information into a final grade. Some teachers did not provide much detail regarding how the composite was formed. For example, one teacher said she "considers attendance, participation, effort, conduct, and teamwork, and adds to this things such as tests and quizzes" in assigning a final grade. Another teacher was more specific about some of the details. She designs the test she uses herself, because:

> "Most of the students usually don't know enough of the material covered in the provided tests to get a passing grade." This teacher uses "an average of 16-20 grades during the grading period in calculating the final grade. However, the lower grades are not factored into the average." To this mix, she adds her "overall impressions of effort and how the class performed."

The practice described by the teacher above is apparently not uncommon. Several teachers reported a highly similar practice, involving throwing out the worst quiz score per

student, considering class performance as a whole, and considering impressions of a student's effort and ability. As an aside, it is worth mentioning that the practice of "throwing out" one or more poor scores on formal assessments is apparently quite widespread; teachers use the practice ostensibly so that an inappropriately low score does not inappropriately affect a student's grade. No teacher reported throwing out an outlier in the other direction--an inappropriately high score that might inappropriately inflate a student's grade, though such a practice would be entirely consistent with the logic of throwing out a low score. We believe this again points to a strong "success orientation" in teachers.

Finally, a few of the teachers made specific mention of using "extra credit" as one component taken into account when assigning the final. Unfortunately, we do not have any details about what sorts of activities fall under this rubric, nor do we have more precise information about the relative contribution of that component.

Professional collaboration

The third theme that was apparent in the teachers' comments was the lack of professional collaboration on matters of assessment. Perhaps many teachers are more active in discussing, planning, interpreting, and coordinating assessment activities; however, the ones in this sample who commented on aspects of collaboration provided a fairly uniform picture of independence. This independence is possibly abetted by the finding reported earlier that many teachers do not share a common knowledge or understanding of district policies on grading, where such policies exist.

The most striking example of isolation in assessment matter came from interviews

with two teachers, a math teacher and a physical education teacher in a secondary school. In commenting on the question that asked about the number of marks they use in calculating final grades compared to what other teachers do, both said that they were unsure about what other teachers do. What made this example so striking was the fact that these two teachers both reported having 25 years of experience in education, and they taught in the same building.

One teacher, who commented earlier about her philosophy that "shaping the kids' minds ... is more important than averaging tests and quiz scores" also said that she realizes her philosophy "is unorthodoxed [sic] and is a bit nervous to share this with the administration or other teachers." Other teachers reported that knew a formal policy for grading existed, but they choose to ignore them. We suspect this occurs because there no mechanisms and forums--or only weak ones--exist for teachers to collaborate on assessment matters or to discuss, formulate, or influence district policies.

Discussion and Recommendations

As others have noted, teachers and administrators often enter the field of teaching without systematic training in assessment. Indeed, it also seems that many teachers need not acquire expertise in classroom assessment to complete advanced certification or retain licensure (O'Sullivan & Chalnick, 1991; Schafer & Lissitz, 1987; Ward, 1980). This study confirms the generally acknowledged weaknesses in the preservice and inservice preparation of teachers in classroom assessment and grading, and that additional assistance in meeting the Standards for Teacher Competence in Educational Assessment of Students

(AFT/NCME/NEA, 1990) is needed. Perhaps the most revealing overall finding is our general inability to discover strong predictors of differential assessment practice. It is often claimed that teachers acquire much of what they know about assessment "on the job" or through "trial and error." One of our most interesting findings was that, for teachers who have not had formal training in testing and grading, very few of the assessment practices we studied were found to be related to years of experience in the profession.

Another stream of research has generally concluded that teachers' assessment practices do not necessarily conform to what measurement specialists would consider to be sound testing and grading practice (see, for example, Stiggins, Frisbie, & Griswold, 1989). For the teachers studied in this research, assessment practices varied widely and unpredictably, with no apparent relationship to characteristics such as practice level, years of experience, gender, or familiarity with district policies. For example, although these teachers reported giving and using an average of approximately 21 marks when calculating students' final grades for a marking period, the variability was quite large (s = 17.9), revealing marked diversity in assessment practices not strongly related to practice setting, gender, or years of experience.

While this variability is undoubtedly attributable somewhat to differences between content areas, it also suggests some cause for concern. These data mean that a large percentage of teachers use fewer than three grades per marking period to arrive at a final grade, while an equally large percentage appear to count nearly everything that comes over the transom. One obvious recommendation would seem to be that many teachers should receive additional assistance in learning how to determine the quality and dependability of information yielded by classroom assessments, formal or otherwise.

We were also surprised by the large percentage of teachers who reported that commercial sources were their primary source for major tests and quizzes. With the increasing appeal of assessment alternatives such as portfolios and performance assessments across the United States, we might have suspected that the number of teachers relying on sources outside the classroom for assessments would be less than the 39% who reported that commercial publishers were their primary source of assessment instruments. This finding warrants further investigation. For example, we wondered: Do teachers have ample time to consider and construct assessments that are carefully integrated with their instruction, or is the reliance on outside sources a matter of expedience? What kinds of changes need to take place to foster teacher involvement in classroom assessment? Even if teachers continue to rely on outside sources to a substantial degree, it would seem that the ability to carefully evaluate, choose, and interpret assessments developed outside the classroom would be an important skill. These observations would seem to provide further support for encouraging relevant training in classroom assessment.

Another surprising finding in this area was that it was beginning teachers (i.e., those with between one and five years of experience) who were most likely to develop their own tests and quizzes, while more experienced teachers were more likely to rely on commercially prepared assessments. Given the pressures beginning teachers face, we would have suspected the opposite.

Variation in the frequency of graded assignments or tests, and reliance on commercially prepared assessments are not necessary major causes for concern. Far more important is the accuracy and relevance of the information gathered and interpretability of

grades as indicators of performance. On these fronts also, this research has provided discomforting data. Teachers generally claim to consider a variety of diverse factors in assigning grades. For example, teachers reported that final grades are often a combination of formal measures of achievement, informal achievement-related information, informal measures of non-achievement related factors, and the kitchen sink. In short, grades appear to consist of a potpourri of elements that vary from district to district, from teacher to teacher within a district, and even from student to student within a classroom.

One fairly consistent finding in the teachers' comments revealed what might be called a "success bias." With discernable regularity, teachers appeared to structure their assessment practices and combine formal and informal assessment information in ways that were most likely to result in a higher grade for their students. This same phenomenon might also help explain why teachers reported including so many factors when assigning final grades.

Beyond this consistency, it is not at all clear that any interested group--administrators, teachers, parents, or even students and teachers themselves--can confidently glean the meaning of the grades students receive. Apparently, the marks that students receive on individual assignments are composites formed by combining a number of elements weighted in idiosyncratic proportions. These marks are then combined to form a final grade, again taking various sources of information into account and combining that information in unknown, but varying ways.

One finding that also may help to explain some of the other results of this study is the large percentage of teachers who reported that their school districts do not have a formal grading policy. A sensible recommendation would seem to be that many districts should

begin to consider, establish, and disseminate information that would provide guidance to teachers about desirable assessment and grading practices. Beyond this, however, it is noteworthy that--even in districts that reportedly have a formal grading policy--a majority of the teachers surveyed indicated that they are unaware of or deliberately ignore those policies. The finding that teachers may be unaware of a macro-level policy may not be all that surprising, given ubiquitous bureaucratic inefficiencies in disseminating information. However, in this study, the teachers surveyed also reported that they are also generally unaware of their colleagues' practices. Teachers interviewed for this study candidly admitted that they ignored district grading policies; several who acknowledged that they were unsure about what their colleagues did vis a vis assessment and grading, also indicated that they preferred it that way. A recommendation that follows from these observations would seem to be that schools more actively pursue engendering cultures of collaborative reflective practice, especially related to assessment. We hope that additional research and professional development efforts are directed toward this goal.

Finally, we note that we anticipate conducting further research in the area of assessment practices. We envision that a comparative replication of this study in which the sample consists of teachers who have had graduate preparation in assessment may provide interesting insights into the effectiveness (or ineffectiveness) of such training on teachers' assessment practices.

Table 1 - Sample Characteristics

Variable		Number	(Percent)
Gender			
	Male	31	(21.7)
	Female	112	(78.3)
Age			
	21-25	11	(7.7)
	26-30	21	(14.7)
	31-35	19	(13.3)
	36-40	14	(9.8)
	41-45	33	(23.1)
	46-50	22	(15.4)
	51-55	16	(11.2)
	56-above	4	(2.8)
	Missing	3	(2.1)
Setting			
	Elementary school	83	(58.0)
	Middle/Junior high school	27	(18.9)
	High school	33	(23.1)
School Location			
	Rural	21	(14.7)
	Suburban	48	(33.6)
	Urban	58	(40.6)
	Other/Missing	16	(11.2)
Teaching Experience			
	1-5 years	39	(27.3)
	6-10 years	26	(18.2)
	11-15 years	26	(18.2)
	16+ years	52	(36.4)

Table 2a - Frequency and Sources of Assessments by Setting

	Setting			
	Elementary	Middle School	High School	Total
Question: How often do you give minor assignments that count for a grade?				
At least once/wk	72.6%	81.5%	75.8%	75.2%
Less than once/wk	27.4%	18.5%	24.2%	24.8%
n	73	27	33	133
Question: How often do you give major tests or exams that count for a grade?				
At least once every two weeks	48.6%	59.3%	60.6%	53.8%
Less frequently than every two weeks	51.4%	40.7%	39.4%	46.2%
n	72	27	33	132
Question: What is your primary source for minor tests and/or quizzes?				
Develop them myself	76.2%	87.5%	84.6%	80.5%
Use publishers' tests	23.8%	12.5%	15.4%	19.5%
n	63	24	26	113
Question: What is your primary source for major tests and/or exams?				
Develop them myself	50.0%	72.7%	74.2%	61.3%*
Use publishers' tests	50.0%	27.3%	25.8%	38.7%
n	58	22	31	111

* $p < .05$ ** $p < .01$

Table 2b - Frequency and Sources of Assessments by Experience

	Years of Experience				
	1-5	6-10	11-15	16+	Total
Question: How often do you give minor assignments that count for a grade?					
At least once/wk	71.1%	68.2%	86.4%	76.5%	75.2%
Less than once/wk	28.4%	31.8%	13.6%	23.5%	24.8%
n	38	22	33	51	133
Question: How often do you give major tests or exams that count for a grade?					
At least once every two weeks	51.4%	54.5%	52.2%	56.0%	53.8%
Less frequently than every two weeks	48.6%	45.5%	47.8%	44.0%	46.2%
n	37	22	23	50	132
Question: What is your primary source for minor tests and/or quizzes?					
Develop them myself	96.9%	78.4%	65.0%	76.2%	80.5%*
Use publishers' tests	3.1%	21.1%	35.0%	23.8%	19.5%
n	32	19	20	42	113
Question: What is your primary source for major tests and/or exams?					
Develop them myself	65.5%	64.7%	73.7%	52.2%	61.3%
Use publishers' tests	34.5%	35.3%	26.3%	47.8%	38.7%
n	29	17	19	46	111

* $p < .05$ ** $p < .01$

Table 2c - Frequency and Sources of Assessments by Gender

	Gender		
	Male	Female	Total
Question: How often do you give minor assignments that count for a grade?			
At least once/wk	64.5%	78.4%	75.2%
Less than once/wk	35.5%	21.6%	24.8%
n	31	102	133
Question: How often do you give major tests or exams that count for a grade?			
At least once every two weeks	48.4%	55.4%	53.8%
Less frequently than every two weeks	51.6%	44.5%	46.2%
n	31	101	132
Question: What is your primary source for minor tests and/or quizzes?			
Develop them myself	83.3%	79.8%	80.5%
Use publishers' tests	16.7%	20.2%	19.5%
n	24	89	113
Question: What is your primary source for major tests and/or exams?			
Develop them myself	81.5%	54.8%	61.3% *
Use publishers' tests	18.5%	45.2%	38.7%
n	27	84	111

* $p < .05$ ** $p < .01$

Table 3a - Factors Considered in Assigning Grades by Setting

	Setting			
	Elementary	Middle School	High School	Total
Question: What factors do you consider when assigning grades on assignments, tests, etc.? (Table values are percentages responding "Yes.")				
Percent or number correct	82.9%	88.9%	81.8%	83.8%
Difficulty of the test	36.8%	33.3%	33.3%	35.3%
How class performed	35.5%	55.6%	51.5%	43.4%
Individual student ability	57.9%	37.0%	48.5%	51.5%
Individual student effort	43.4%	40.7%	39.4%	41.9%
n	76	27	33	136
Question: What sources of information do you use to assign final grades for a marking period? (Percent responding "Yes.")				
Formal achievement measures (e.g., tests, assignments)	88.2%	88.9%	90.9%	89.0%
Other formal measures (e.g., attendance, class participation, etc.)	43.4%	55.6%	69.7%	52.2% *
Informal achievement measures (e.g., students' answers to in-class and individual questions)	44.7%	40.7%	36.4%	41.9%
Other informal measures (e.g., impressions of effort, conduct, teamwork)	68.4%	55.6%	48.5%	61.0%
n	76	27	33	136

Table 3a (continued)

	Setting			
	Elementary	Middle School	High School	Total
Question: What does the final grade represent? (Percent responding "Yes.)				
Final grade represents...				
Individual student achievement on fixed classroom goals.	36.6%	33.0%	32.3%	34.9%
Individual student achievement on fixed goals, but considering overall class performance.	15.5%	18.5%	25.8%	18.6%
Individual student achievement on individualized goals.	19.7%	14.8%	16.1%	17.8%
Combination of group and individual achievement on fixed classroom goals.	28.2%	33.3%	25.8%	28.7%
n	71	27	31	129

* $p < .05$ ** $p < .01$

Table 3b - Factors Considered in Assigning Grades by Experience

	Years of Experience				
	1-5	6-10	11-15	16+	Total
Question: What factors do you consider when assigning grades on assignments, tests, etc.? (Table values are percentages responding "Yes.")					
Percent or number correct	94.7%	73.9%	79.2%	82.4%	83.8%
Difficulty of the test	42.1%	26.1%	33.3%	35.3%	35.3%
How class performed	44.7%	39.1%	37.5%	47.1%	43.4%
Individual student ability	44.7%	56.5%	41.7%	58.8%	51.5%
Individual student effort	36.8%	47.8%	29.2%	49.0%	41.9%
n	38	23	24	51	136
Question: What sources of information do you use to assign final grades for a marking period? (Percent responding "Yes.")					
Formal achievement measures (e.g., tests, assignments)	89.5%	95.7%	87.5%	86.3%	89.0%
Other formal measures (e.g., attendance, class participation, etc.)	44.7%	47.8%	66.7%	52.9%	52.2%
Informal achievement measures (e.g., students' answers to in-class and individual questions)	31.6%	30.4%	50.0%	51.0%	41.9%
Other informal measures (e.g., impressions of effort, conduct, teamwork)	63.2%	47.8%	58.3%	66.7%	61.0%
n	38	23	24	51	136

Table 3b (continued)

	Years of Experience				
	1-5	6-10	11-15	16+	Total
Question: What does the final grade represent? (Percent responding "Yes.)					
Final grade represents...					
Individual student achievement on fixed classroom goals.	36.1%	26.1%	40.9%	35.4%	34.9%
Individual student achievement on fixed goals, but considering overall class performance.	19.4%	39.1%	9.1%	12.5%	18.6%
Individual student achievement on individualized goals.	13.9%	13.0%	31.8%	16.7%	17.8%
Combination of group and individual achievement on fixed classroom goals.	30.6%	21.7%	18.2%	35.4%	28.7%
n	36	23	22	48	129

* $p < .05$ ** $p < .01$

Table 3c - Factors Considered in Assigning Grades by Gender

	Gender		
	Male	Female	Total
Question: What factors do you consider when assigning grades on assignments, tests, etc.? (Table values are percentages responding "Yes.")			
Percent or number correct	87.1%	82.9%	83.8%
Difficulty of the test	32.3%	36.2%	35.3%
How class performed	38.7%	44.8%	43.4%
Individual student ability	48.4%	52.4%	51.5%
Individual student effort	48.4%	40.0%	41.9%
n	31	105	136
Question: What sources of information do you use to assign final grades for a marking period? (Percent responding "Yes.")			
Formal achievement measures (e.g., tests, assignments)	87.1%	89.5%	89.0%
Other formal measures (e.g., attendance, class participation, etc.)	54.8%	51.4%	52.2%
Informal achievement measures (e.g., students' answers to in-class and individual questions)	29.0%	45.7%	41.9%
Other informal measures (e.g., impressions of effort, conduct, teamwork)	51.6%	63.8%	61.0%
n	31	105	136

Table 3c (continued)

Gender

	Male	Female	Total
Question: What does the final grade represent? (Percent responding "Yes.)			
Final grade represents...			
Individual student achievement on fixed classroom goals.	38.7%	33.7%	34.9%
Individual student achievement on fixed goals, but considering overall class performance.	16.1%	19.4%	18.6%
Individual student achievement on individualized goals.	16.1%	19.4%	17.8%
Combination of group and individual achievement on fixed classroom goals.	29.0%	28.6%	28.7%
n	31	98	129

* $p < .05$ ** $p < .01$

Table 4a - Grades and Grading Policies by Setting

	Setting			
	Elementary	Middle School	High School	Total
Question: How many grades are usually included when calculating each student's final grade?				
Mean (standard deviation)	21.5 (14.9)	30.4 (19.8)	24.5 (20.6)	24.3 (17.9)
n	62	27	31	120
Question: How does the total number of grades you use in calculating students' final grades compare to that of most other teachers in your building?				
Most use fewer grades	11.6%	25.9%	24.2%	17.8%
Most use about the same	29.0%	29.6%	9.1%	24.0%
Most use more	10.1%	0.0%	0.0%	5.4%
Not sure	49.3%	44.4%	66.7%	52.7%
n	69	27	33	129
Question: Does your district have a formal grading policy?				
Yes	50.7%	63.0%	60.6%	55.6%
No	30.7%	33.3%	36.4%	32.6%
Not sure	18.7%	3.7%	3.0%	11.9%
n	75	27	33	135

* $p < .05$ ** $p < .01$

Table 4b - Grades and Grading Policies by Experience

	Years of Experience				
	1-5	6-10	11-15	16+	Total
Question: How many grades are usually included when calculating each student's final grade?					
Mean (standard deviation)	21.6 (14.2)	30.4 (23.1)	23.0 (13.1)	24.0 (19.0)	24.3 (17.6)
n	36	22	17	45	120
Question: How does the total number of grades you use in calculating students' final grades compare to that of most other teachers in your building?					
Most use fewer grades	11.4%	21.7%	22.7%	18.4%	17.8%
Most use about the same	31.4%	26.1%	22.7%	18.4%	24.0%
Most use more	11.4%	4.3%	0.0%	4.1%	5.4%
Not sure	45.7%	47.8%	54.5%	59.2%	52.7%
n	35	23	22	49	129
Question: Does your district have a formal grading policy?					
Yes	48.6%	44.0%	52.2%	68.0%	55.6%
No	32.4%	40.0%	39.1%	26.0%	32.6%
Not sure	18.9%	16.0%	8.7%	6.0%	11.9%
n	37	25	23	50	135

* $p < .05$ ** $p < .01$

Table 4c - Grades and Grading Policies by Gender

	Gender		
	Male	Female	Total
Question: How many grades are usually included when calculating each student's final grade?			
Mean (standard deviation)	22.2 (18.4)	24.9 (17.8)	24.3 (17.9)
n	28	92	120
Question: How does the total number of grades you use in calculating students' final grades compare to that of most other teachers in your building?			
Most use fewer grades	19.4%	17.3%	17.8%
Most use about the same	22.6%	24.5%	24.0%
Most use more	0.0	7.1%	5.4%
Not sure	58.1%	51.0%	52.7%
n	31	98	129
Question: Does your district have a formal grading policy?			
Yes	61.3%	53.8%	55.6%
No	35.5%	31.7%	32.6%
Not sure	3.2%	14.4%	11.9%
n	31	104	135

* $p < .05$ ** $p < .01$

References

Airasian, P. W. (1994). Classroom assessment, (2nd ed.). New York: McGraw-Hill.

Airasian, P. W. (1991). Perspectives on measurement instruction. Educational Measurement: Issues and Practices, 10(1), 13-16, 26.

American Federation of Teachers, National Council on Measurement in Education, National Education Association (1990). Standards for teacher competence in educational assessment of students. Washington, DC: Author.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. Journal of Educational Measurement, 30(2), 123-142.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. Applied Measurement in Education, 7(4), 279-301.

Cizek, G. J., and Rachor, R. E. (1994, October). Classroom grading practices: Who's doing what, and why? Paper presented at the annual meeting of the Midwestern Educational Research Association, Chicago, IL.

Durm, M. W. (1993). An A is not an A is not an A: A history of grading. Educational Forum, 57, 294-297.

Friedman, S. J. & Frisbie, D. A. (1993, April). The validity of report cards as indicators of student performance. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Gorney, B. E. (1994, April). An integrative analysis of the content and structure of school report cards: What school systems report to the public. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Griswold, P. A. & Griswold, M. M. (1992, April). The grading contingency: Graders' beliefs and expectations about the assessment ingredients. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Gullickson, A. R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23(4), 347-354.

Hills, J. R. (1991). Apathy concerning testing and grading. Phi Delta Kappan, 72(7), 540-545.

Hoge, R. D. & Coladarci, T. (1989). Teacher based judgments of academic achievement: A review of literature. Review of Educational Research, 59(3), 297-313.

Jones, M. S. (1990, April). Preservice teachers' beliefs about grading practices. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Manke, M. P. & Loyd, B. (1991, April). A study of teachers' understanding of their grading practices. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Nava, F. J. G. & Loyd, B. H. (1992, April). The effect of student characteristics on the grading process. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

O'Sullivan, R. G. & Chalnick, M. K. (1991). Measurement-related course work requirements for teacher certification and recertification. Educational Measurement: Issues and Practice, 10(1), 17-19, 23.

Schafer, W. D. & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38(3), 57-63.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high school work in English. School Review, 20(3), 442-457.

Starch, D., & Elliott, E. C. (1913a). Reliability of grading work in history. School Review, 21(2), 254-295.

Starch, D., & Elliott, E. C. (1913b). Reliability of grading work in history. School Review, 21(4), 676-681.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. Phi Delta Kappan, 69(5), 363-368.

Stiggins, R. J. (1991a). Assessment literacy. Phi Delta Kappan, 72(10), 534-539.

Stiggins, R. J. & Bridgeford, J. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271-286.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. Educational Measurement: Issues and Practice, 8(2), 5-14.

Stiggins, R. J. (1991b). Relevant classroom assessment training for teachers. Educational Measurement: Issues and Practice, 10(1), 7-12.

Stiggins, R. J. & Conklin, N. F. (1988). Teacher training in assessment. Portland, OR: Northwest Regional Education Laboratory.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. Educational Measurement: Issues and Practice, 8(2), 5-14.

Terwilliger, J. S. (1989). Classroom standard setting and grading practices. Educational Measurement: Issues and Practice, 8(2), 15-19.

Waltman, K. K. & Frisbie, D. A. (1993). Parents' understanding of their children's report card grades. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Ward, J. G. (1980). Teachers and testing: A survey of knowledge and attitudes. In L.M. Rudner (Ed.), Testing in our schools, (pp. 15-24). Washington, DC: National Institute of Education.

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. Journal of Teacher Education, 42(1), 37-42.

Wood, P., Bennett, T., Wood, J., & Bennett, C. (1990, April). Grading and evaluation practices and policies of school teachers. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA.